Localization of auditory sources spatialized in virtual environments

Localização de fontes auditivas espacializadas em ambientes virtuais

Laura Berenice Piceda
Laboratorio de Acústica y Percepción Sonora (LAPSo), Universidad Nacional de Quilmes, CONICET, R. S. Peña 352, Bernal, Buenos Aires, B1876BXD, Argentina
laura.piceda@unq.edu.ar

Esteban N. Lombera
Universidad Nacional de Tres de Febrero (UNTREF), Departamento de Ciencia y Tecnología. Sáenz Peña, Argentina
elombera@untref.edu.ar

Ramiro Vergara
Laboratorio de Acústica y Percepión Sonora (LAPSo), Universidad Nacional de Quilmes, CONICET, R. S. Peña 352, Bernal, Buenos Aires, B1876BXD, Argentina
ramiro.vergara@unq.edu.ar

Abstract: This study aims to investigate the correlation between the spatial location of virtual sound sources, as simulated by the freely available plug-in Facebook Spatial Workstation 360, and the perceived location by a group of listeners. Although sound spatialization techniques based on sound source simulation are widely utilized in the field of music, few studies have systematically evaluated their perceptual effectiveness, and limitations remain unexplored. Therefore, this study seeks to obtain psychoacoustic experiment-based data that would not only improve digital sound spatialization techniques but also facilitate a better understanding of the limits of perception associated with them. To achieve this objective, three experiments were designed, each focusing on a single variable (i.e., azimuth, elevation, and distance). The findings of this study suggest that the plugin effectively simulates the acoustic environment similarly to what has been reported in previous studies using non-individualized virtual stimuli, meaning it manages to emulate angular origin in azimuth and auditory distance to a considerable extent while failing to emulate the elevation of sound sources. Additionally, our results showed a high rate of front-back and back-front confusion.

Keywords: audition. perception. psychoacoustics. auditory localization. virtual reality.

Resumo: Este estudo tem como objetivo investigar a correlação entre a localização espacial das fontes sonoras virtuais, conforme simulada pelo plug-in gratuito Facebook Spatial Workstation 360, e a localização percebida por um grupo de ouvintes. Embora as técnicas de espacialização sonora baseadas na simulação de fontes sonoras sejam amplamente utilizadas no campo da música, poucos estudos avaliaram sistematicamente sua eficácia perceptual, e limitações ainda permanecem inexploradas. Portanto, este estudo busca obter dados baseados em experimentos psicoacústicos que não apenas melhorem as técnicas de espacialização sonora digital, mas também facilitem uma melhor compreensão dos limites de percepção associados a elas. Para atingir esse objetivo, foram projetados três experimentos, cada um focando em uma variável específica (ou seja, azimute, elevação e distância). Os resultados deste estudo sugerem que o plug-in simula efetivamente o ambiente acústico de maneira semelhante ao que foi relatado em estudos anteriores utilizando estímulos virtuais não individualizados, o que significa que ele consegue emular a origem angular no azimute e a distância auditiva de forma considerável, enquanto falha em emular a elevação das fontes sonoras. Além disso, nossos resultados mostraram uma alta taxa de confusão entre frente e trás e entre trás e frente.

Palavras-chave: audição. Percepção. Psicoacústica. localização auditiva. realidade virtual.

Submitted on: 11 January, 2025
Approved on: 8 March, 2025
Published on: April, 2025

1. Introduction

Spatial hearing, or the ability to locate sounds, is a fundamental function of the auditory system that allows individuals to determine both the direction and distance of sounds in their environment (Middlebrooks, 2015). This ability is critical in everyday life, enabling us to locate potential threats, navigate our surroundings, and communicate effectively with others. Additionally, spatial hearing is essential in the entertainment industry, particularly in music production and performance, where it plays a crucial role in creating an immersive and engaging experience for listeners.

The brain must recognize three fundamental factors in order to perceive the location of a sound source: the angular origin of the sound in azimuth and elevation and the distance to the source.

The human auditory system excels in horizontal localization, primarily relying on interaural differences in time (ITD) and intensity (ILD) cues (Blauert 1997; Sodnik, Bobojevic, Susnik and Tomazic, 2004). ITD, influenced by sound source location and interaural distance, aids in determining azimuthal direction, remarkably accurate for low-frequency sounds (below 1.5 kHz) (Letowski and Letowski, 2012; Middlebrooks, 2015). ILD, stemming from partial occlusion of sound waves by the head and body, results in intensity differences between ears, more pronounced for high-frequency sounds (above 4 kHz), while low-frequency sounds produce minimal ILD regardless of location (Letowski and Letowski, 2012; Middlebrooks, 2015).

To perceive the elevation of a sound source and its position relative to the head, humans rely on monaural spectral cues. Changes in the sound spectrum generate these cues as they interact with the listener's body, mainly the head, ears, and torso, before reaching the eardrum. The resulting signal that reaches the eardrum is the original sound filtered by its interaction with the listener's body (Algazi, Avendano and Duda, 2001; Bloom, 1977). The filtering process that the signal undergoes as it interacts with the body is known as the head-related transfer function (HRTF)

(Middlebrooks, 1999; Li and Peissig, 2020). HRTFs are unique for each individual as they depend on the position of the sound source and the listener's physiognomy. Therefore, personalized HRTFs are necessary to simulate spatialized sound accurately for a specific listener. HRTFs exhibit peaks and valleys at different locations in the frequency spectrum, enabling elevation estimation through psychophysical and acoustic studies (Shimoda, Nakashima, Kumon, Kohzawa, Mizumoto and Iwai, 2006).

The brain employs a range of cues to perceive the distance of a sound source, with sound intensity being the most influential cue. As the distance from the source increases, the value of sound intensity decreases. Another important cue is the energy ratio between direct and reverberant sound (D/R). In reverberant environments, the intensity of the direct sound decreases by 6 dB for each doubling of distance, while the energy of the reflected portion remains relatively constant. In addition, auditory distance perception is influenced by cues such as spectral content, source familiarity, binaural cues, and visual cues related to the source and environment (Kolarik, Cirstea, Pardhan and Moore, 2016)

1.1 Spatial Hearing in Music

Humans have long used auditory space in art, especially music. Early civilizations enhanced music in reverberant spaces like caves and temples. In classical Greek theater (5th-3rd centuries B.C.), masks amplified actors' voices (Knudsen, 1932). During the Middle Ages, church architecture was designed for acoustic effects in sacred music. Renaissance composers like Giovanni Gabrieli (1557-1612) experimented with spatial sound using multiple choirs and ensembles (Bono, 2007). The systematic use of sound space began in the early 20th century with technological advances in speaker arrangements (Abregú, Calcagno, and Vergara, 2012). Spatial sound has since been integral to stereo, quadraphonic systems, and 5.1 surround formats in cinema, TV, and home audio (Anderson et al., 2009).

Although these sound spatialization systems can simulate sound spatiality, they are limited in simulating 360º sound source locations. Advanced techniques like Ambisonics, binaural methods, and sophisticated panning laws have been developed to address this. These methods have gained popularity with immersive virtual technology, providing powerful simulation tools to the public since they offer a more immersive listening experience by accurately representing spatial information, creating a greater sense of depth and dimension in sound. These systems also improve playback quality in various acoustic environments by allowing precise control over sound placement. Initially confined to labs, these technologies are now widely available through software like Blue Ripple Sound 03A Core, Dolby Atmos VR, Facebook Spatial Workstation 360, and others (Hong, Lam, Gupta and Gan, 2017; Gould, 2018; Munoz 2020; Pitman, 2021).

The application of these technologies within music-related contexts encompasses a broad spectrum of domains. These include Extended Reality (XR) (Munoz, 2020), Virtual Reality music composition (Janer, Gomez, Martorell, Miron and Wit, 2016; Serafin, Erkut, Kojs, Nilsson and Nordahl, 2016; Reed and Phelps, 2019; Pitman, 2021), immersive video game music (Webster, Garnier and Sedes, 2017; Pitman, 2021), electroacoustic music (Webster et al., 2017), and soundscapes and live performances (Vicuña Zubiria, 2018; Malyshev, 2018; Obadiah, 2023). Furthermore, these technologies find application across various musical styles that seek to evoke a sensation of being enveloped or surrounded by sound (Lynch and Sazdov, 2017; Rada Hurtado, 2019).

However, few systematic studies have focused on validating these simulations with respect to the actual location perceived by listeners. Scientific validation of open-source virtual auditory environments can benefit both artistic and scientific domains by facilitating discourse among users from various disciplines and aiding developers in creating scientifically validated simulations. Therefore, validating these environments can help to ensure their accuracy and efficacy in reproducing a realistic auditory scene.

This paper focuses on investigating the correlation between virtual auditory sources and their perceived spatial location using the Facebook Spatial Workstation 360º plug-in. This free suite supports various formats, including binaural, ambisonics, stereo, and VR/AR (Hong, 2017). Comprehensive documentation aids implementation in music contexts. Key features include a spatializer plug-in with object movement icons on a video window, easy head-locked stereo via the "Stereo Master" bus and FB360 Encoder, live playback through Oculus Rift DK2, and a standalone encoder for multiple output formats (France, 2017; Pavoni, 2021). The plug-in also supports head-locked stereo channels, enabling distinct sound stages for elements like music and voice-over, enhancing its use in artistic contexts (Reed and Phelps, 2019).

It is pertinent to note that the support for the Facebook Audio 360° Spatial Workstation ceased subsequent to the conclusion of our experiments. However, its relevance remains due to its user base, ongoing projects, and the absence of comparable alternatives. The existence of an active Facebook group dedicated to this software, characterized by a substantial (more than nine thousand members at the time of writing this paper) online community (France, 2017; Reed and Phelps, 2019) which continues to expand presently, along with user-generated resources facilitating access, serves to underscore its sustained significance within the sphere of audio production. Furthermore, its widespread adoption among industry professionals, alongside scholarly assessments of its attributes (Gupta et al, 2017; Malyshev, 2018; Vicuña Zubiria, 2018; Rada Hurtado, 2019; Reed and Phelps, 2019; Rana, Ozcinar and Smolic 2019; Pavoni, 2021; Obadiah, 2023) emphasizes the necessity of scrutinizing its functionality and role within the domain of spatial audio tools.

In the current study, we conducted psychoacoustic experiments to evaluate the spatial localization of sound sources simulated using Facebook 360, focusing on azimuth, elevation, and distance. This will allow us to establish a baseline for users to understand what to expect from the plug-in, which is easier when its limitations are known.

2. Method

The method used for the experiments is subsequently described:

2.1 Participants

Eighteen volunteers (7 women and 11 men) between 18 and 60 years old (mean age = 33 years) with no declared hearing pathologies participated in the experiment. Participants were recruited through Facebook advertisements and e-mails. All the participants performed the experiment at home using their personal computers. This condition is similar to what actually occurs with this type of commercial development, as it is used in varied environments and with diverse equipment.

All participants provided written informed consent before the start of the experiment. No data sets were deleted.

2.2 Signals

The stimulus employed across all three experiments of this study was a 500 ms white noise burst featuring a 20 ms fade-in and fade-out. This decision was made considering that various authors have reported that the precision of absolute localization improves with the expansion of signal bandwidth (Burger, 1958; Butler, 1986; Middlebrooks, 1992; Blauert, 1997; Carlile, Leong, and Hyams, 1997; King and Oldfield, 1997; Best et al., 2005).

Binaural room impulse responses (BRIRs), were generated and spatialized using the Facebook 360 Spatializer plug-in, a tool designed for binaural audio rendering within immersive environments. Operating within an Ambisonic framework, the plug-in encodes sound sources in a full-sphere spatial field using three-dimensional Cartesian coordinates (X, Y, Z). Its core functionality includes HRTF processing, real-time spatial positioning, and Ambisonic rendering, simulating how sound waves interact with the listener's head and ears to create a sense of spatial depth and directionality.

To ensure accurate spatialization, each stimulus was assigned to an independent audio channel and routed to a dedicated spatial audio bus, where the plug-in applied directional filtering and binaural processing. The spatialized audio was then encoded in an Ambisonic format and converted for binaural playback.

The virtual acoustic environment was configured as a rectangular room measuring 18m x 20m x 12m (width x length x height). However, since the Facebook 360 Spatializer plug-in documentation does not provide explicit details on the reverb characteristics of the virtual space, an impulse response measurement technique was employed using a logarithmic sine sweep (Farina, 2000). The recorded response was deconvolved with the inverse filter to extract the room's impulse response, allowing for the estimation of the reverberation time (T60 = 0.322 s). This additional analysis provided insight into the spatial characteristics imposed by the plug-in's rendering process, ensuring a better understanding of its impact on auditory perception.

2.3 Design

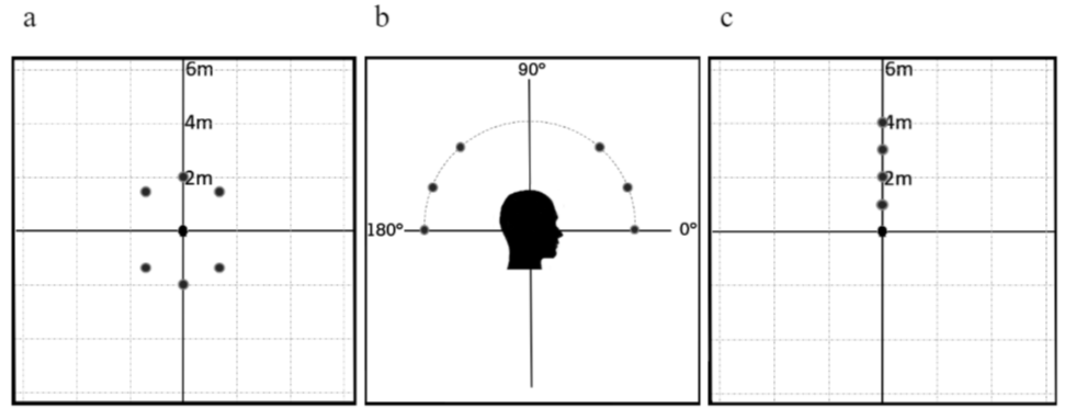
The following psychoacoustic experiments were conducted:

Experiment 1 -Azimuth: Six different audios were used, each repeated five times for each location randomly. The stimuli were placed at the same distance (2 m) and elevation (0°) at the following azimuth positions: 0° (front), 45°, -45°, 135°, -135° and 180° (back), with negative values being those located to the left of the subject (Figure 1a).

Experiment 2 -Elevation: The experiment consisted of six different audios, each repeated five times for each location randomly. The stimuli were placed at the same horizontal angle (azimuth 0º) and a fixed distance (2 m) at the following elevations:0º (front), 20º, 50º, 130º, 160º and 180º (back) (Figure 1b).

Experiment 3 -Distance: Four different audios were incorporated, each repeated five times for each distance randomly. The generated stimuli were placed at the same horizontal angle (azimuth 0º) and elevation (0º) at the following distances: 1m, 2m, 3m, and 4m (Figure 1c).

Figure 1: Graphic representation of source locations in a) Azimuth; b) Elevation; and c) Distance.



2.4 Procedure

This study employed a response method similar to Kopčo and Shinn-Cunningham (2011). With a metric grid, participants viewed the surrounding space from above and used a response interface to indicate stimulus locations freely. To ensure consistency in auditory spatialization, participants were required to always use headphones. Before each experiment, a chime played on the left and another on the right, serving as a verification step for correct headphone placement. Participants adjusted the volume to a comfortable level and replayed the chimes if they had to remove or reposition the headphones during the session.

To maintain data reliability, the experiment had to be conducted in a quiet indoor environment free from external noise and distractions. The instructions stated this requirement, and participants were asked to confirm compliance before proceeding. Since the study relied on participants' home environments, complex interfaces or calibrated monitors were not always feasible, necessitating standard mouse or keyboard responses (Eerola, Armitage, Lavan, and Knight, 2021).

The experiment was supervised remotely via real-time video calls. The experimenter remained online throughout each session to address participant inquiries, provide clarifications, and ensure proper procedure execution. Participants were required to share their screens during the experiment, allowing the experimenter to monitor their progress and verify adherence to the instructions. After completing the experiments, participants submitted their responses via a file transfer website (Lavan, Burston, and Garrido, 2019; Njie, Lavan, and McGettigan, 2021).

2.5 Data Analysis

To analyse the results, a Python script automatically extracted the central pixel's position relative to the coordinate origin and converted it into degrees or meters based on predefined scales. The resulting data table was imported into R and integrated with other tables detailing stimuli, order, and repetition sequence. Graphical representations included individual and population averages, variability (standard deviation), and estimation precision (standard error). Unsigned absolute bias (UB) measured estimation error regardless of direction. Intra-subject comparisons were employed for statistical analysis at $\alpha = 0.05$ significance level. Each experiment was analyzed separately before integration to assess software spatialization quality.

2.6 Ethical Considerations

The experiments were performed with the written consent of each participant, following the Code of Ethics of the World Medical Association (Declaration of Helsinki), and were approved by the Ethics Committee of the Universidad Nacional de Quilmes. All participants provided written consent and did not receive any compensation.

3. Results

The results obtained for each experiment are subsequently described:

3.1 Experiment 1- Azimuth

For the frontal stimulus (0°), the bulk of the responses were clustered in the proximity of 0° and 180°. This implies that a significant number of listeners perceived the source at positions opposite to the virtual source location. Something similar occurs when the stimulus is located behind the listener (180°). The remaining stimuli (45°, -45°, 135°, and -135°) were perceived as fully lateralized, that is, tending to collapse around 90° and -90°. The responses corresponding to these stimuli were analyzed by paired comparisons using t-tests applied to the paired difference as a tool. The results of these comparisons are presented in Table 1. Comparisons between stimuli located on the same side (-45° vs. -135º and 45º vs. 135º) yielded no significant differences ($p = 0.29$ and 0.42, respectively), which is indicative of the subject's inability to distinguish whether the stimulus was located in a frontal or posterior position. On the other hand, comparisons between stimuli located in the front (-45º vs 45º) or back (-135º vs 135º) yield significant differences ($p < 0.0001$ in both cases), which is indicative of subjects being able to clearly distinguish the laterality of the stimulus.

Table 1- Paired comparisons for lateral stimuli

Comparisons	Statistic	p
Left: -135º vs -45º	t(17) = 1.09	0.29
Right: 135º vs 45º	t(17) = 0.83	0.42
Front: -45º vs 45º	t(17) = -15.1	<10-10
Back: -135º vs 135º	t(17) = -18.7	<10-12

Front-back//back-front confusion with these stimuli is prevalent, posing a challenge for appropriate treatment. Past studies resolved confusion by coding responses for the correct hemisphere (Wightman and Kistler, 1989b) or excluded them from descriptive statistics (Makous and Middlebrooks, 1990) to prevent error inflation. We excluded front-back//back-front confusions and reported their rates separately within each condition.

Figure 2 shows individual averages for each stimulus: squares denote responses without confusion, and crosses indicate confusion. Population averages (and standard errors), calculated from responses without confusion, are represented by dots. Hexagons denote sound source locations. This presentation facilitates visualization and comparison of individual and population averages.

The tendency to misjudge the front-back direction of the sound source was substantial. When a sound was presented in the front hemisphere, participants incorrectly perceived it as coming from the back 47.76% of the time. Conversely, a sound originating from the back was mistakenly reported as coming from the front in 50.37% of cases. This suggests a nearly symmetrical confusion pattern between front and back locations.

The average perceived location of sound sources across different azimuth angles, along with the standard error of the mean (SEM), is as follows: for a source at 0° (directly in front), the average response was slightly shifted forward at 4.7° ± 6.1°. When the source was positioned at 45° to the right, participants localized it at 59.3° ± 6.5°, indicating a slight but consistent overestimation. Similarly, for -45° (left), the average perceived direction was -64.4° ± 4.3°, again slightly exaggerated. At 180° (directly behind), responses were centered around 176.3° ± 11.4°, showing a slight forward bias. For lateral sources at 135° and -135°, the perceived locations were 114.7° ± 4.5° and -111.1° ± 3.2°, respectively, reflecting mild compression towards the midline.

The overall variability in participants' responses, measured as the standard deviation (SD), averaged ± 23.33°. Some positions exhibited more significant inconsistency: for instance, responses at 180° (back) showed the highest variation (SD = ± 42.7°), likely due to front-back confusion, while lateral positions like -135° (left-back) had the lowest variability (SD = ± 12.6°), suggesting more consistent perception.

To assess individual performance stability, we calculated intra-subject standard deviation (I-S SD), which measures how consistently each participant localized sounds across trials, yielding a mean of ± 13.28°. Additionally, intra-subject unsigned bias (I-S UB), which quantifies absolute localization errors without considering direction, had an average value of 25.71°. These results indicate that while participants generally followed a predictable pattern, individual accuracy varied, and front-back confusions significantly impacted performance.

3.2 Experiment 2- Elevation

Figure 3 represents individual averages: squares denote responses without confusion, and crosses indicate confusion. The population average (and standard error) is shown as a dot-based only on responses without confusion. Hexagons mark sound source locations, facilitating visualization and comparing individual and population averages.

Participants responded uniformly within the permitted range, suggesting difficulty in assigning a definitive elevation direction for each source. Consequently, the average perceived elevation was near 45° (for responses without confusion) or 90° (for all responses), reflecting the center of the response range. Statistical analysis was not applied due to the inadequate representation of the distribution by the mean. Data shows an apparent lack of consistency between virtual source elevation and perceived elevation. Most subjects perceived source elevation over the entire available range of responses for a given source position. The exception to this observation corresponds to subjects who concentrated their responses in a narrow range of elevation, regardless of the actual source position.

Figure 2. Azimuth Responses. This figure presents participants' azimuthal localization responses. The head illustration represents the virtual listener's position. Mean responses (M ± SEM) are displayed as data points with error bars, showing overall response trends. Individual participant averages are distinguished using shape coding: squares indicate responses without front-back confusion, while crosses represent confused responses. The actual sound source locations are marked with hexagons. Additional statistical measures are provided for each position, including the standard deviation (SD) of responses, intra-subject standard deviation (I-S SD) reflecting individual response consistency, intra-subject unsigned bias (I-S UB) quantifying directional accuracy, and the Back-Front//Front-Back Error (F-B/F-B E), which captures instances of front-back perceptual reversals.

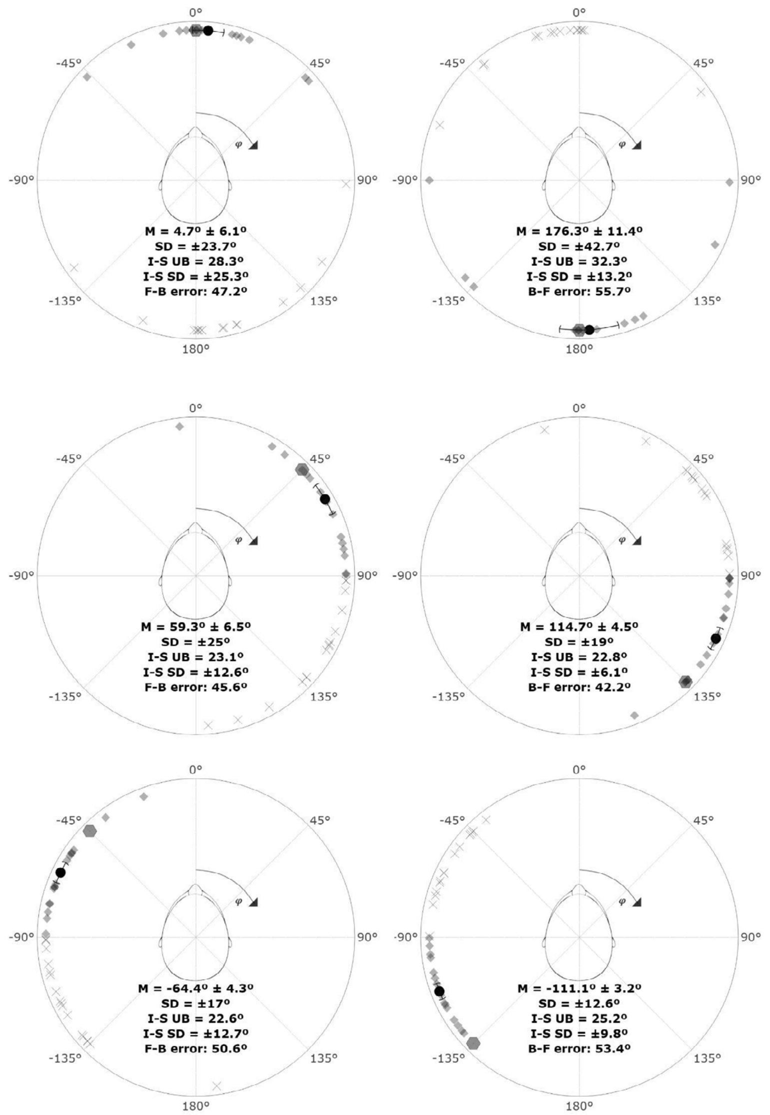
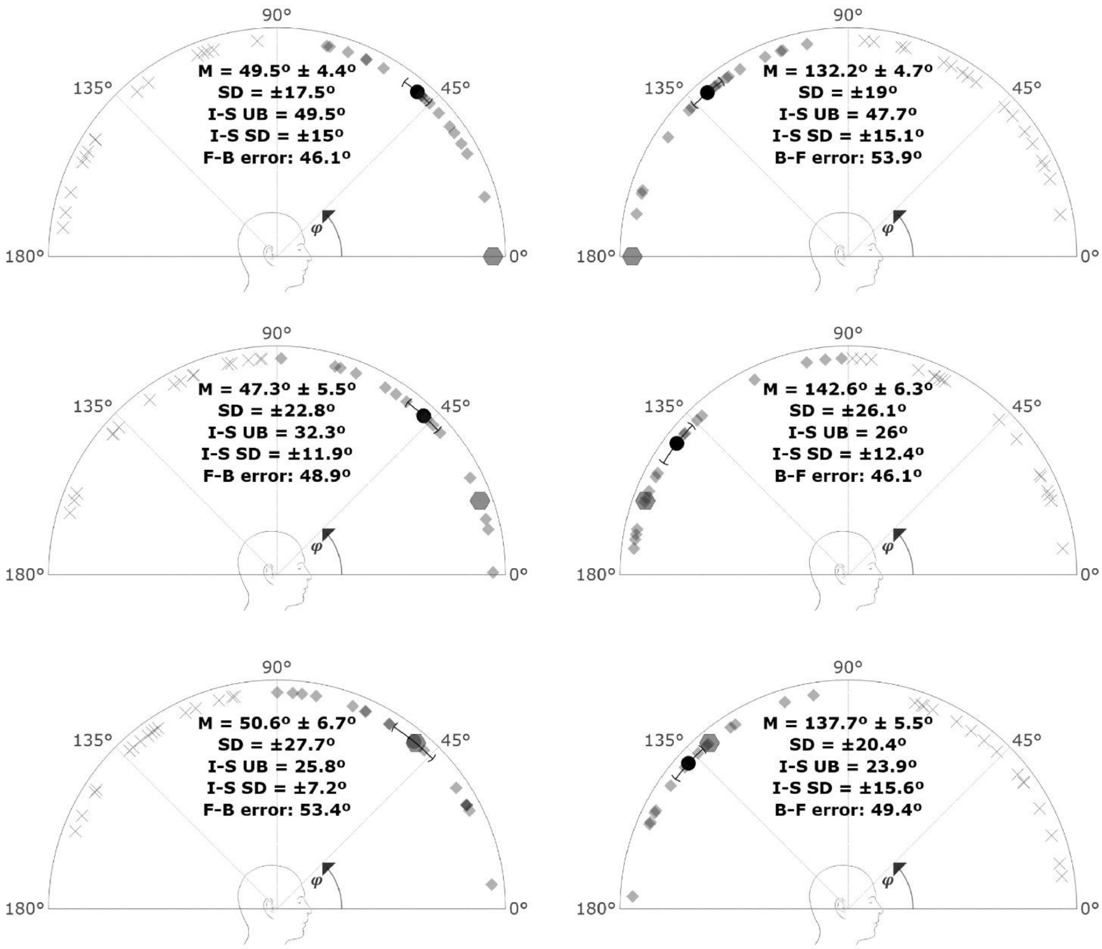


Figure 3. Elevation Responses. This figure illustrates participants' elevation localization responses. The head illustration, viewed from above, represents the virtual listener's position relative to the stimulus presentation. Mean responses (M ± SEM) are shown as data points with error bars to indicate participant variability. Individual response patterns are shape-coded: squares represent responses without confusion, crosses indicate front-back confusions, and hexagons denote the actual sound source positions. The figure also includes key statistical measures: the standard deviation (SD) of mean responses, intra-subject standard deviation (I-S SD) to assess response variability within participants, intra-subject unsigned bias (I-S UB) to measure absolute localization errors, and Back-Front/Front-Back Error (F-B/F-B E) to quantify misperceptions of sound directionality. These metrics provide a comprehensive view of elevation localization performance.



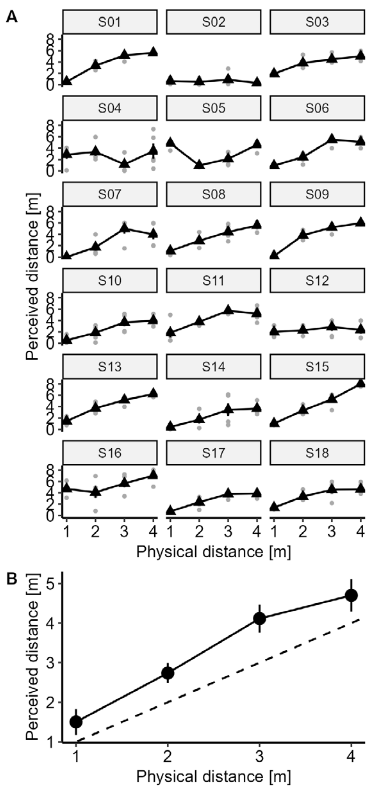
The standard deviations and unsigned biases (intra-subject) corresponding to both elevation and azimuth (considering responses without confusion) were scrutinized through paired comparisons employing t-tests on the paired differences as analytical tools.

While comparisons between stimuli positioned in elevation versus azimuth did not demonstrate significant differences in the results of intra-subject standard deviation (p = 0.91), notable distinctions were observed for the unsigned bias (p = 0.0018).

3.3 Experiment 3- Distance

Figure 4a displays individual responses for each subject: dots represent responses for each presentation, triangles indicate individual averages. Meanwhile, Figure 4b illustrates the perceived distance averaged across subjects. Participants consistently overestimate source locations, aligning with the source's actual range (1.5 to 4.7 m vs. 1 to 4 m). Notably, no participant perceived the source in the posterior region, contrasting with experiment 1 (azimuth). Some participants perceived azimuthal displacement, which was evident in their responses exceeding the grid limit (6m).

Figure 4 a) Perceived distance of physical distance for each subject b) Intrasubject average of perceived distance of physical distance (mean +/- standard error). Note: The dashed line shows perfect response (slope=1, ordinate to origin=0) for reference.



To quantify previous observations, responses underwent analysis via a linear mixed-effects model, incorporating a fixed "source position" effect and subject-specific random slope and ordinate to the origin. A significant effect of source position was detected [F(1,53)=46.7, p < .0001], indicating subjects discerned changes in source distance. Model parameter estimates are detailed in Table 2. Notably, the slope slightly exceeds 1, while the ordinate to the origin approximates 0.5 meters.

Table 2- Parameter estimates for a linear fit over the distance curve

	Estimate	Standard Error	Statistic	p
Ordenate	0.521 m	0.340 m	t(53) = 1.53	0.13
Slopé	1.097	0.161	t(53) = 6.83	<.00001

We calculated the signed relative bias to quantify overestimation, defined as S = (Y - X) / X, where Y is the response and X is the source position. A positive bias (S > 0) indicates overestimation, while a negative bias indicates underestimation. Analysis via a linear model with the same effect structure as perceived distance revealed no significant effect of target position [F(1,53) = 0.914, p = 0.34], suggesting relative bias is a fixed proportion relative to source distance. Subsequently, individual biases were collapsed along source distance for each subject. The intrasubject distribution of collapsed bias was analyzed, confirming its mean significantly differs from 0 (one-sample t-test, t(17) = 3.08, p = 0.0067, M=0.35, 95% CI [0.11, 0.60]), indicating a consistent tendency toward overestimation.

4. Discussion

In this study, we sought to measure the plausibility with which the Facebook Spatial Workstation 360° plug-in simulates the location of static sound sources. To this end, we performed three experiments to measure the perceived location of the simulated sources at different azimuth, elevation, and distance positions. We conducted a comparative analysis between the outcomes garnered from preceding studies conducted in both real and virtual environments, encompassing both individualized and non-individualized. This examination aimed to elucidate the perceptual capabilities and limitations inherent in the simulations executed in this study.

Even in real environments, auditory localization precision and fidelity face noise and uncertainty, causing deviations in sound source perception (Blauert, 1983). Studies show that virtual sound source localization is generally less precise than in real-world settings. However, similar performance levels are reported when virtual environments use personalized HRTFs. Since commercial virtual environments typically lack individualized simulations, their fidelity and precision are expected to be lower than those of real-world or customized virtual sources (see Letowski and Letowski 2012 for a detailed review).

Various techniques are available to improve the fidelity of simulations that utilize non-individualized stimuli. When full individual measurements are not feasible, perceptual tuning of generic HRTFs has been proposed to enhance localization accuracy (Middlebrooks, 1999; Romigh et al., 2015). Another viable approach is the adaptive selection from a predefined HRTF database based on perceptual similarity metrics, allowing users to select the best-matching set without the need for complex measurement procedures. These methods provide practical alternatives for implementing individualized HRTFs in virtual spatial audio systems, making them more applicable in both commercial and research settings.

Another possibility for improving non-individualized stimuli's performance is training tasks with multisensory feedback. This type of feedback has been shown to enhance performance in various auditory localization tasks, both for real (Hüg et al., 2019) and virtual sound sources (Berguer et al., 2018). For example, Parise et al. (2024) showed that sensorimotor training rapidly induces adaptation to altered spectral cues for sound localization. In just 20 minutes, participants fully adapted to modified cues, with improvements in localization gain, bias, and precision. These results highlight the short-term plasticity of human spatial hearing, which quickly adjusts to spectral changes. These findings suggest that the performance of non-individualized virtual environments could be enhanced if platforms offered multisensory feedback training alternatives.

In summary, the fidelity of spatial audio in virtual environments plays a critical role in creating immersive experiences, particularly in VR, where auditory and visual cues collaborate to strengthen the sense of presence (Slater et al., 2009). In music applications, especially those involving 3D audio for immersive listening, spatial inaccuracies can disrupt depth and positioning perception, diminishing the overall experience (Blauert, 1997).

Thus, future research should prioritize developing more precise methods for customizing virtual stimuli to enhance realism and immersion across various domains.

4.1 Perceptual Evaluation

The results of the angular localization experiments revealed significant confusion between front and back orientations. In experiment 1 (azimuth localization), there was a front-back confusion rate of 46.3% and a back-front rate of 50.3%. In experiment 2 (elevation localization), both confusion rates were around 49.4%-49.8%. This suggests that about half of the participants misperceived the source's location, compared to real-world error rates of 2%-12% (Hollander 1994; Oldfield and Parker 1986; Wightman and Kistler 1989; Makous and Middlebrooks 1989; Wenzel, Arruda, Kistler, and Wightman 1993). Virtual sources show higher confusion rates: 12%-20% with individualized BRIRs and 15%-52% with non-individualized BRIRs (Begault and Wenzel 1993; Besing and Koehnke 1995; Bronkhorst 1995; Wenzel et al. 1993). Our findings, with confusion rates around 50%, align with these elevated rates. Schonstein, Ferré, and Katz (2008) reported 37.5%-52% error rates for non-individualized sources, Wenzel et al. 1993) reported 20%-58%. Our results showed similar rates, differing from studies showing more front-back inversions (Chasin and Chong 2008, 1998). Factors like sound source visibility, auditory environment, sound spectrum, and headphone type affect these rates (Letowski and Letowski 2012), complicating interpretation due to varied experimental conditions.

Experiment 1 (azimuth perception) showed biases of 22° to 30° with individual standard deviations of 10° to 25°. This indicates reduced precision compared to controlled lab settings, where errors range from 2° to 15° (Best et al., 2009; Carlile et al. 1997; Makous and Middlebrooks, 1990; Tiitinen, Palomäki, Mäkinen, May and Alku, 2004), and real-world errors of 3° to 6° (Carlile et al., 1997). In virtual environments, horizontal localization errors range from 15° to over 30° (Wenzel and Foster, 1993; Middlebrooks, 1999; Begault, Wenzel and Anderson, 2001; Endsley and Rosiles, 1995), matching our findings. Most listeners can derive azimuth cues using non-individualized HRTFs (Wenzel et al., 1993). Additionally, listeners overestimated source angles outside the midplane, perceiving them closer to the interaural axis, consistent with studies showing overestimation of lateral positions by 5°–15° in both natural (Oldfield and Parker, 1984a) and virtual environments (Carlile et al., 1997; Majdak, Goupell and Laback, 2010).

Experiment 2 (elevation perception) revealed that none of the participants accurately perceived the source's elevation. Participants showed more significant bias in elevation perception than azimuth, with similar intra-subject variability. Figure 3 indicates a uniform distribution of responses regardless of the source's actual location. These results align with Mendonça, Campos, Dias, Vieira, Ferreira and Santos (2012), where participants using non-individualized broadband stimuli couldn't discern elevation. Some participants perceived sources without elevation (0° and 180°) as higher, consistent with prior research (Folds, 2006; Mendonça et al., 2012). This issue arises from not considering individual anatomical characteristics. (Wenzel et al., 1993) showed that localization errors decrease with individually measured HRTFs. In real-world scenarios, perceived elevation can be distorted by up to 21.9° when outer ear folds are obstructed (Oldfield and Parker, 1984, 1986; Hofman, 1998). In summary, the Facebook Spatial Workstation 360º plug-in failed to accurately simulate sound source elevation, likely due to its non-individualized simulations.

In experiment 3 (distance perception), listeners showed a linear response with consistent overestimation across all distances. The obtained response was striking, as it was more veridical than those previously reported under similar conditions (Kolarik, Moore, Zahorik, Cirstea and Pardhan, 2016). We posit that factors associated with the response methodology may have influenced listeners' responses. In our experiments, participants mapped their perceived distances on a scaled grid spreadsheet, with options extending up to a maximum of 6m. Although the specific test distances were not delineated, participants had indirect visual cues regarding the simulated room, such as its maximum size. Prior studies (Calcagno, Abregu, Eguía and Vergara, 2012; Kolarik, Pardhan, Cirstea and Moore, 2013) have demonstrated that visual information pertaining to a room can enhance distance perception. Similar responses were observed in studies with virtual sound sources and visual contextual information (Cabrera and Gilfillan, 2002; Cubick, 2015). This suggests that distance perception can be influenced by indirect contextual cues, such as room dimensions conveyed through floor plans or diagrams. We deemed it intriguing to investigate this hypothesis further. A noteworthy observation is that, unlike the outcomes observed in the angular perception experiments, all participants in the distance perception experiment perceived the source in front of their heads. Understanding the reason for this discrepancy is challenging, particularly considering that the stimuli employed were identical across experiments. We hypothesize that this result could be attributed to differences in the visual examples used to elucidate the experiment for participants. The examples provided for distance perception showcased sources exclusively in front of the listeners' heads, whereas the examples for the azimuth and elevation angular perception experiments depicted sources located both in front and behind. It is plausible that this influenced participants' responses. It would be interesting to test this hypothesis in future research.

One limitation of our study is the reliance on participants using their own computers and headphones in home environments, which may introduce variability compared to controlled laboratory settings (Schonstein et al., 2008). Differences in headphone models and quality could particularly affect auditory cues related to timbre and spectral characteristics, which are crucial for the perception of elevation and distance. However, azimuthal localization is expected to be less impacted, as ITD and ILD are relatively robust in terms of variations in headphone quality.

Despite this potential variability, this methodology provides valuable insights into real-world application performance across diverse environments and equipment, enhancing the study's ecological validity (Hong et al., 2017). Consistency in spatial perception across these varied settings suggests robustness in the auditory processing mechanisms involved (Eerola et al., 2021). Furthermore, spatial audio technologies are generally designed for use with a range of conventional consumer headphones, making this variability a relevant aspect of real-world listening scenarios.

A key limitation is that we did not systematically record information about the specific headphones used by participants. This omission represents an opportunity for future research to explore how different domestic configurations, particularly variations in headphone quality and frequency response, affect 3D audio reproduction. Given the limited number of studies directly comparing localization performance across different headphone models, investigating this factor in future experiments could provide valuable insights into optimizing spatial audio experiences for diverse user setups (Anglada-Tort, Harrison & Jacoby, 2022; Pfordresher & Demorest, 2021; Harrison et al., 2020).

5. Conclusions

Our findings suggest that the Facebook Spatial Workstation 360º plug-in is valuable for spatial perception, particularly in azimuth and auditory distance. However, it lacks fidelity in simulating sound source elevation, with a high rate of front-back and back-front confusion, consistent with prior research in non-individualized virtual environments. Alternative methodologies, such as individualized HRTF-based environments, may be necessary for precise and tailored virtual auditory stimuli. Nonetheless, the plug-in's ability to foster spatial perception makes it useful for specific users, especially in audiovisual and musical arts. Additionally, the plug-in is designed to complement visual stimuli, which can enhance localization accuracy (King, 2009). The synchronous presentation of spatially disparate visual and auditory stimuli may induce the Ventriloquist effect, which aligns the perceived sound source with the visual stimulus location (Bertelson and Radeau, 1981). Thus, in audiovisual environments, both "localization blur" and "front-back and back-front confusions" may approximate real-world experiences. Future investigations could explore the influence of visual cues on sound-source localization accuracy when using this plug-in. Overall, while the Facebook Spatial Workstation 360º exhibits the typical constraints of non-individualized virtual environments, its accessibility and utility make it valuable for crafting immersive audiovisual and musical experiences.

References

ABREGÚ, E. L.; CALCAGNO, E. R.; VERGARA, R. La distancia como dimensión estructural en la música. **Revista Argentina de Musicología**, v. 12-13, p. 379-400, 2012.

ALGAZI, V. R.; AVENDANO, C.; DUDA, R. O. Elevation localization and head-related transfer function analysis at low frequencies. **The Journal of the Acoustical Society of America**, v. 109, n. 3, p. 1110-1122, 2001. DOI: https://doi.org/10.1121/1.1349185

ALGAZI, V. R.; DUDA, R. O.; THOMPSON, D. M.; AVENDANO, C. The cipic hrtf database. In: **Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)**, 2001, p. 99-102. DOI: https://doi.org/10.1109/ASPAA.2001.969552

ANDERSON, J. et al. Música y espacio: ciencia, tecnología y estética. Buenos Aires: **Universidad Nacional de Quilmes**, 2009.

ANGLADA-TORT, M.; HARRISON, P. M.; JACOBY, N. REPP: A robust cross-platform solution for online sensorimotor synchronization experiments. **Behavior Research Methods**, v. 54, n. 5, p. 2271-2285, 2022.

BEGAULT, D. R.; WENZEL, E. M. Headphone localization of speech. **Human Factors**, v. 35, n. 2, p. 361-376, 1993. DOI: https://doi.org/10.1177/001872089303500210

BEGAULT, D. R.; WENZEL, E. M.; ANDERSON, M. R. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. **Journal of the Audio Engineering Society**, v. 49, n. 10, p. 904-916, 2001.

BERGER, C. C.; GONZALEZ-FRANCO, M.; TAJADURA-JIMÉNEZ, A.; FLORENCIO, D.; ZHANG, Z. Generic HRTFs may be good enough in virtual reality: Improving source localization through cross-modal plasticity. **Frontiers in Neuroscience**, v. 12, p. 21. https://doi.org/10.3389/fnins.2018.00021

BERTELSON, P.; ASCHERSLEBEN, G. Temporal ventriloquism: crossmodal interaction on the time dimension: 1. evidence from auditory–visual temporal order judgment. **International Journal of Psychophysiology**, v. 50, n. 1-2, p. 147-155, 2003.

BESING, J. M.; KOEHNKE, J. A test of virtual auditory localization. **Ear and Hearing**, v. 16, n. 2, p. 220-229, 1995.

BEST, V. et al. The role of high frequencies in speech localization. **The Journal of the Acoustical Society of America**, v. 118, n. 1, p. 353-363, 2005.

BEST, V. et al. A meta-analysis of localization errors made in the anechoic free field. In: **Principles and applications of spatial hearing**, p. 14-23, 2011. DOI: https://doi.org/10.1142/9789814299312_0002

BLAUERT, J. Spatial hearing: The psychophysics of human sound localization. Cambridge: **MIT Press**, 1997.

BLOOM, P. J. Determination of monaural sensitivity changes due to the pinna by use of minimumaudiblefield measurements in the lateral vertical plane. **The Journal of the Acoustical Society of America**, v. 61, n. 3, p. 820-828, 1977. DOI: https://doi.org/10.1121/1.381346

BRONKHORST, A. W. Localization of real and virtual sound sources. **The Journal of the Acoustical Society of America**, v. 98, n. 5, p. 2542-2553, 1995. DOI: https://doi.org/10.1121/1.413219

BRONKHORST, A. W. The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. **Acta Acustica united with Acustica**, v. 86, n. 1, p. 117-128, 2000.

BURGER, J. F. Front-back discrimination of the hearing systems. **Acta Acustica United with Acustica**, v. 8, n. 5, p. 301-302, 1958.

BUTLER, R. A. The bandwidth effect on monaural and binaural localization. **Hearing Research**, v. 21, n. 1, p. 67-73, 1986.

CABRERA, D.; GILFILLAN, D. Auditory distance perception of speech in the presence of noise. **Georgia Institute of Technology**, 2002.

CALCAGNO, E. R.; ABREGÚ, E. L.; EGUÍA, M. C.; VERGARA, R. The role of vision in auditory distance perception. **Perception**, v. 41, n. 2, p. 175-192, 2012. DOI: https://doi.org/10.1068/p7153

CARLILE, S.; LEONG, P.; HYAMS, S. The nature and distribution of errors in sound localization by human listeners. **Hearing Research**, v. 114, n. 1-2, p. 179-196, 1997. DOI: https://doi.org/10.1016/S0378-5955(97)00161-5

CUBICK, J.; SANTURETTE, S.; LAUGESEN, S.; DAU, T. The influence of visual cues on auditory distance perception. **Forschritte der Akustik DAGA'15**, p. 1220-1223, 2015.

EEROLA, T.; ARMITAGE, J.; LAVAN, N.; KNIGHT, S. Online data collection in auditory perception and cognition research: Recruitment, testing, data quality and ethical considerations. **Auditory Perception & Cognition**, v. 4, n. 3-4, p. 251-280, 2021.

ENDSLEY, M. R.; ROSILES, S. A. Auditory localization for spatial orientation. **Journal of Vestibular Research: Equilibrium & Orientation**, v. 5, n. 6, p. 473-485, 1995.

FARINA, A. Simultaneous measurement of impulse response and distortion with a swept-sine technique. **Audio Engineering Society Convention 108**, 2000.

FOLDS, D. J. The elevation illusion in virtual audio. In: **Proceedings of the Human Factors and Ergonomics Society Annual Meeting**, v. 50, n. 16, p. 1576-1579. Los Angeles, CA: **Sage CA**, 2006. https://doi.org/10.1177/154193120605001613

FRANCE, M. Immersive Audio Production: Providing structure to research and development in an emerging production format. Doctoral dissertation, **Ohio University**, 2017.

GOULD, R.; LANE, C. Let's test: 3d audio spatialization plugins. **Designing Sound: The Art & Technique of Sound Design**, 2018.

GUPTA, R.; LAM, B.; HONG, J.; ONG, Z. T.; GAN, W. S.; CHONG, S. H.; FENG, J. 3D audio AR/VR capture and reproduction setup for auralization of soundscapes. In: **Proceedings of the 24th Intl. Congress on Sound and Vibration, ICSV24**, 2017.

HARRISON, P.; MARJIEH, R.; ADOLFI, F.; VAN RIJN, P.; ANGLADA-TORT, M.; TCHERNICHOVSKI, O.; et al. Gibbs sampling with people. **Advances in neural information processing systems**, v. 33, p. 10659-10671, 2020.

HOFMAN, P. M.; VAN RISWICK, J. G.; VAN OPSTAL, A. J. Relearning sound localization with new ears. **Nature Neuroscience**, v. 1, n. 5, p. 417-421, 1998. https://doi.org/10.1038/1633

HOLLANDER, A. J. An exploration of virtual auditory shape perception. Master's thesis, **University of Washington**, 1994.

HONG, J. Y.; HE, J.; LAM, B.; GUPTA, R.; GAN, W. S. Spatial audio for soundscape design: Recording and reproduction. **Applied Sciences**, v. 7, n. 6, p. 627, 2017.

HÜG, M. X.; VERGARA, R. O.; TOMMASINI, F. C.; ETCHEMENDY, P. E.; BERMEJO, F.; FERNANDEZ, L. G. Reaching measures and feedback effects in auditory peripersonal space. **Scientific Reports**, v. 9, n. 1, p. 9476, 2019. https://doi.org/10.1038/s41598-019-45912-4

JANER, J.; GOMEZ, E.; MARTORELL, A.; MIRON, M.; DE WIT, B. Immersive orchestras: audio processing for orchestral music VR content. In: **2016 8th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)**, 2016, p. 1-2. **IEEE**.

KIDD Jr, G.; ARBOGAST, T. L.; MASON, C. R.; GALLUN, F. J. The advantage of knowing where to listen. **The Journal of the Acoustical Society of America**, v. 118, n. 6, p. 3804-3815, 2005. https://doi.org/10.1121/1.2109187

KING, R. B.; OLDFIELD, S. R. The impact of signal bandwidth on auditory localization: Implications for the design of three-dimensional audio displays. **Human Factors**, v. 39, n. 2, p. 287-295, 1997. https://doi.org/10.1518/001872097778543895

KING, A. J. Visual influences on auditory spatial learning. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 364, n. 1515, p. 331-339, 2009. https://doi.org/10.1098/rstb.2008.0230

KNUDSEN, V. O. Recent developments in architectural acoustics. **Reviews of Modern Physics**, v. 6, n. 1, p. 1, 1934. https://doi.org/10.1103/RevModPhys.6.1

KOLARIK, A. J.; PARDHAN, S.; CIRSTEA, S.; MOORE, B. C. Using acoustic information to perceive room size: effects of blindness, room reverberation time, and stimulus. **Perception**, v. 42, n. 9, p. 985-990, 2013. https://doi.org/10.1068/p7555

KOLARIK, A. J.; MOORE, B. C.; ZAHORIK, P.; CIRSTEA, S.; PARDHAN, S. Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. **Attention, Perception, & Psychophysics**, v. 78, n. 2, p. 373-395, 2016. https://link.springer.com/article/10.3758/s13414-015-1015-1

KOPČO, N.; SHINN-CUNNINGHAM, B. G. Effect of stimulus spectrum on distance perception for nearby sources. **The Journal of the Acoustical Society of America**, v. 130, n. 3, p. 1530-1541, 2011. https://doi.org/10.1121/1.3613705

LAVAN, N.; BURSTON, L. F.; GARRIDO, L. How many voices did you hear? Natural variability disrupts identity perception from unfamiliar voices. **British Journal of Psychology**, v. 110, n. 3, p. 576-593, 2019. https://doi.org/10.1111/bjop.12348

LETOWSKI, T. R.; LETOWSKI, S. T. Auditory spatial perception: Auditory localization. **Army Research Lab Aberdeen Proving Ground MD Human Research and Engineering Directorate**, 2012.

LI, S.; PEISSIG, J. Measurement of head-related transfer functions: A review. **Applied Sciences**, v. 10, n. 14, p. 5014, 2020. https://doi.org/10.3390/app10145014

LYNCH, H.; SAZDOV, R. A perceptual investigation into spatialization techniques used in multichannel electroacoustic music for envelopment and engulfment. **Computer Music Journal**, v. 41, n. 1, p. 13-33, 2017.

MAKOUS, J. C.; MIDDLEBROOKS, J. C. Two-dimensional sound localization by human listeners. **The Journal of the Acoustical Society of America**, v. 87, n. 5, p. 2188-2200, 1990. https://doi.org/10.1121/1.399186

MALYSHEV, M. Sound production for 360 videos: in a live music performance case study. 2018.

MAJDAK, P.; GOUPELL, M. J.; LABACK, B. 3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training. **Attention, Perception, & Psychophysics**, v. 72, n. 2, p. 454-469, 2010. https://doi.org/10.3758/APP.72.2.454

MENDONÇA, C.; CAMPOS, G.; DIAS, P.; VIEIRA, J.; FERREIRA, J. P.; SANTOS, J. A. On the improvement of localization accuracy with non-individualized HRTF-based sounds. **Journal of the Audio Engineering Society**, v. 60, n. 10, p. 821-830, 2012.

MIDDLEBROOKS, J. C.; MAKOUS, J. C.; GREEN, D. M. Directional sensitivity of sound-pressure levels in the human ear canal. **The Journal of the Acoustical Society of America**, v. 86, n. 1, p. 89-108, 1989. https://doi.org/10.1121/1.398224

MIDDLEBROOKS, J. C. Narrow-band sound localization related to external ear acoustics. **The Journal of the Acoustical Society of America**, v. 92, n. 5, p. 2607-2624, 1992.

MIDDLEBROOKS, J. C. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. **The Journal of the Acoustical Society of America**, v. 106, n. 3, p. 1493-1510, 1999. https://doi.org/10.1121/1.427147

MIDDLEBROOKS, J. C. Sound localization. **Handbook of Clinical Neurology**, v. 129, p. 99-116, 2015.

MIRALLES BONO, J. L. El espacio como recurso musical. 2011.

MOORE, B. C. An introduction to the psychology of hearing. **Brill**, 2012.

MUÑOZ, I. G. Space Time Exploration of Musical Instruments. **University of California, San Diego**, 2020.

NEISSER, U. Cognition and Reality. San Francisco: **Freeman**, 1976.

NJIE, S.; LAVAN, N.; MCGETTIGAN, C. Talker and accent familiarity yield advantages for voice identity perception: A voice sorting study. **PsyArXiv**, 2021. https://doi.org/10.31234/osf.io/b6ftg

OLDFIELD, S. R.; PARKER, S. P. Acuity of sound localisation: A topography of auditory space. I. Normal hearing conditions. **Perception**, v. 13, n. 5, p. 581-600, 1984. https://doi.org/10.1068/p130581

OLDFIELD, S. R.; PARKER, S. P. Acuity of sound localisation: A topography of auditory space. III. Monaural hearing conditions. **Perception**, v. 15, n. 1, p. 67-81, 1986.

PARISE, C.; GORI, M.; FINOCCHIETTI, S.; ERNST, M.; ESPOSITO, D.; TONELLI, A. Happy new ears: rapid adaptation to novel spectral cues in vertical sound localization. **iScience**, v. 27, n. 12, 2024. https://doi.org/10.1016/j.isci.2024.108415

PAVONI, R. Turn Your Head and Listen: 360° Audio Between Old Utopias and Market Strategies. **Cinergie–Il Cinema e le altre Arti**, v. 19, p. 157-169, 2021.

PFDRESHER, P. Q.; DEMOREST, S. M. The prevalence and correlates of accurate singing. **Journal of Research in Music Education**, v. 69, n. 1, p. 5-23, 2021. https://doi.org/10.1177/0022429420951630

PITMAN, D. Sounds in Space: A portfolio of spatial music works for VR. Doctoral dissertation, 2021.

RADA HURTADO, C. A. Audio inmersivo para conciertos silenciosos. Doctoral dissertation, **Universidad de las Artes**, 2019.

RANA, A.; OZCINAR, C.; SMOLIC, A. Towards generating ambisonics using audio-visual cue for virtual reality. In: **ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**, 2019, p. 2012-2016. **IEEE**.

REED, L.; PHELPS, P. Audio reproduction in virtual reality cinemas–Position paper. In: **2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)**, 2019, p. 1513-1516. **IEEE**.

ROMIGH, G. D.; BRUNGART, D. S.; STERN, R. M.; SIMPSON, B. D. Efficient real spherical harmonic representation of head-related transfer functions. **IEEE Journal of Selected Topics in Signal Processing**, v. 9, n. 5, p. 921-930, 2015.

SAVEL, S. Individual differences and left/right asymmetries in auditory space perception. I. Localization of low-frequency sounds in free field. **Hearing Research**, v. 255, n. 1-2, p. 142-154, 2009. https://doi.org/10.1016/j.heares.2009.06.013

SCHONSTEIN, D.; FERRÉ, L.; KATZ, B. F. Comparison of headphones and equalization for virtual auditory source localization. **Journal of the Acoustical Society of America**, v. 123, n. 5, p. 3724, 2008.

SERAFIN, S.; ERKUT, C.; KOJS, J.; NILSSON, N. C.; NORDAHL, R. Virtual reality musical instruments: State of the art, design principles, and future directions. **Computer Music Journal**, v. 40, n. 3, p. 22-40, 2016.

SHIMODA, T.; NAKASHIMA, T.; KUMON, M.; KOHZAWA, R.; MIZUMOTO, I.; IWAI, Z. Spectral cues for robust sound localization with pinnae. In: **2006 IEEE/RSJ International Conference on Intelligent Robots and Systems**, 2006, p. 386-391. **IEEE**. https://doi.org/10.1109/IROS.2006.281827

SLATER, M.; LOTTO, B.; ARNOLD, M. M.; SANCHEZ-VIVES, M. V. How we experience immersive virtual environments: the concept of presence and its measurement. **Anuario de Psicología**, v. 40, n. 2, p. 193-210, 2009.

SODNIK, J.; BOBOJEVIC, G.; SUSNIK, R.; TOMAZIC, S. The directional resolution of virtual sound sources in humans. **Electrotechnical Review**, v. 71, n. 3, p. 121-127, 2004.

TIITINEN, H.; PALOMÄKI, K. J.; MÄKINEN, V.; MAY, P. J.; ALKU, P. Behavioral detection of spatial stimuli is reflected in auditory cortical dynamics. **Neurology & Clinical Neurophysiology: NCN**, 2004, p. 50-50.

WEBSTER, C.; GARNIER, F.; SEDES, A. Empty Room, how to compose and spatialize electroacoustic music in VR in ambisonic and binaural. **International Journal of Virtual Reality**, v. 17, n. 2, 2017.

WENZEL, E. M.; ARRUDA, M.; KISTLER, D.; WIGHTMAN, F. Localization using non-individualized head-related transfer functions. **Journal of the Acoustical Society of America**, v. 94, p. 111-123, 1993.

WENZEL, E. M.; FOSTER, S. H. Perceptual consequences of interpolating head-related transfer functions during spatial synthesis. In: **Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics**, 1993, p. 102-105. **IEEE**. https://doi.org/10.1109/ASPAA.1993.379986

WIGHTMAN, F. L.; KISTLER, D. J. Headphone simulation of free-field listening. I: stimulus synthesis. **The Journal of the Acoustical Society of America**, v. 85, n. 2, p. 858-867, 1989. https://doi.org/10.1121/1.397557

WIGHTMAN, F. L.; KISTLER, D. J. Headphone simulation of free-field listening. II: Psychophysical validation. **The Journal of the Acoustical Society of America**, v. 85, n. 2, p. 868-878, 1989. https://doi.org/10.1121/1.397558

Research ethics committee approval

The experiments were performed with the written consent of each participant, following the Code of Ethics of the World Medical Association (Declaration of Helsinki), and were approved by the Ethics Committee of the Universidad Nacional de Quilmes. All participants provided written consent and did not receive any compensation.

Publisher

Federal University of Goiás. School of Music and Performing Arts. Graduate Program in Music. Publication in the Portal of Periodicals UFG.

The ideas expressed in this article are the responsibility of their authors and do not necessarily represent the opinion of the editors or the university.